National Child Care Information Center

A service of the Child Care Bureau

NCCIC

10530 Rosehaven Street, Suite 400
Fairfax, Virginia 22030
Phone: (800) 616-2242 Fax: (800) 716-2242 TTY: (800) 516-2242
World Wide Web: http://nccic.org

ASSESSMENT AND EVALUATION:
BECOMING AN EDUCATED CONSUMER

PART II: PROGRAM EVALUATION

This document is intended to provide basic information, examples or definitions, and resources on program evaluation. An understanding of these issues and where to go for further information and expertise will assist policy-makers as they participate in State discussions around school readiness, program evaluation, and accountability to ensure the quality of early care and education and to promote the school success of all young children. This resource will help you become an "educated consumer" on this topic. There are two more documents in this series— Part I addresses child assessment, and Part III addresses accountability systems.

Note: The term assessment can be used for a variety of purposes, such as needs assessment, or assessment of property taxes, etc. In the field, assessment most commonly refers to child assessment, and evaluation is most often associated with programs

WHAT IS PROGRAM EVALUATION?

Program evaluation is a systematic process of clearly articulating the services and desired outcomes of an intervention or program. Data tied to the program goals serve to document the quality and effectiveness of the program over time. Program evaluation can incorporate many methods of data collection, such as child assessment, observations of practice, measures of the environment, and surveys or interviews of teachers or parents.

KEY POINTS:

➢ Effective program evaluation depends on the clear identification of (1) the goals and intended outcomes of the program; (2) the services/intervention expected to achieve the outcomes; (3) the amount or level of intervention expected to cause effects; and (4) the relevant data and its collection and analysis.

➢ A commonly used approach, the *Logic Model*, takes you through this process in a step-by-step manner.

➢ It is critically important to be sure that the program has been implemented well and that the logic model—or how you intend to directly influence the expected outcomes—has been clearly defined before measuring impact on programs.

➢ When measuring the impact on children, it is important that the goals and outcomes have a logical relationship to children's development and that children have received the services for a sufficient length of time to allow for a significant impact—given the myriad of factors that affect how children develop.

➢ Choose measures that clearly match the purpose of the evaluation. Data gathered on individual children, when aggregated, can be used to make determinations about the efficacy of the program to help children achieve goals, assuming there is an appropriate link between the program goals and the assessment tool.

➢ Sampling is an effective method to determine child outcomes, while reducing staff burden and costs of individually assessing many children.

➢ Program evaluation results have many important uses for improving the quality of services to children and families, garnering public support, policy development, and justifying public funding.

DEFINITIONS OF TERMS ASSOCIATED WITH PROGRAM EVALUATION:

What does *Aggregate* mean? To *aggregate* data is to take all the individual scores and combine them.

What is *Formative Evaluation*? *Formative evaluation* is the ongoing documentation of the program and its impact that is used to make adjustments and revisions to services or the intermediate goals of the program during the initial phases of program implementation.

What is a *Logic Model*? A *logic model* is a clear and logical explanation of how the goals of the program link to the services provided and produce the expected results.

What is *Sampling*? *Sampling* is a process of selecting a subgroup of a population that will be used to represent the entire population.

What does *Statistically Significant* mean? A result that is reported as *statistically significant* is one that has a very high probability of not occurring by chance.

What is *Summative Evaluation*? *Summative evaluation* is the final summary of the program and its impact at the end of a particular length of time, and it typically measures the long-term goals or impact of the program.

RESOURCES:

For a discussion of the reasons to conduct evaluation, key question answered in evaluations, and the logic model approach, see the following:

■ *Evaluating Early Childhood Programs: Improving Quality and Informing Policy* (July 2003), Vol. 23, No.6, by Walter Gilliam and Valerie Leiter, produced by ZERO TO THREE. This resource is available on the Web at http://zerotothree.org/ztt_journal.html.

For detailed information on how to develop logic models for programs, with many examples and checklists, see the following:

▪ *Logic Model Development Guide: Using Logic Models to Bring Together Planning, Evaluation, and Action* (January 2004), produced by the W.K. Kellogg Foundation. This resource is available on the Web at http://www.wkkf.org/Programming/ResourceOverview.aspx?CID=281&ID=3669.

For information on how to hire an outside evaluator, questions for interviewing evaluators, and how to design and manage an evaluation contract, see the following:

▪ *When and How to Use External Evaluators* (November 2002), by Tracy Rutnik and Marty Campbell, produced by the Annie E. Casey Foundation through the Funders Evaluation Initiative. This resource is available on the Web at http://www.aecf.org/publications/data/using_external_evaluators.pdf.

For a discussion of current challenges in early childhood evaluation, see the following:

▪ "Evaluating Early Childhood Services: What's Really Behind the Curtain" (Summer 2004), in *The Evaluation Exchange*, Volume X, No. 2, by Jack Shonkoff. This resource is available on the Web at http://www.gse.harvard.edu/hfrp/eval/issue26/index.html.

For a detailed discussion, including frequently asked questions and developmental charts from birth to age 8 on how to effectively assess young children when curriculum, assessment, and program evaluation and accountability are interrelated, see the following:

▪ *Early Childhood Curriculum, Assessment, and Program Evaluation: Building an Effective, Accountable System in Programs for Children Birth through Age 8* (November 2003), a revised Joint Position Statement of the National Association for the Education of Young Children (NAEYC) and the National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE). This resource is available on the Web at http://www.naeyc.org/about/positions/pdf/CAPEexpand.pdf.

For specific recommendations for evaluating programs in a way that is appropriate for young English language learners, see the following:

▪ *Screening and Assessment of Young English-Language Learners: Draft Recommendations* (January 2005), published by NAEYC, was developed by a workgroup of experts in the field in collaboration with NAEYC staff as a supplement to NAEYC's position statement—*Early Childhood Curriculum, Assessment, and Program Evaluation: Building an Effective, Accountable System in Programs for Children Birth through Age 8.* The draft is currently out for public comment and will be revised. This resource is available on the Web at http://www.naeyc.org/statements/pdf/englearners.pdf.

HOW ARE STATES USING PROGRAM EVALUATION?

In an era of increasing emphasis on accountability, States are often required to evaluate programs funded with public and/or private dollars. States often use program evaluation results to advocate for increased funding and to show efficacy to the public, including parents, community leaders, and business partners. States use evaluation data to improve or realign services and target scarce resources. Some States are building internal capacity to conduct evaluations, and many States are partnering with universities or private organizations to conduct evaluations.

KEY POINTS:

➢ Summative program evaluation should not be undertaken before a formative evaluation reveals that the program has been implemented well and of a sufficient intensity in order to reasonably expect effects on child, family, or caregiver practice.

➢ A commitment to invest in the professional development and training necessary for all staff that have responsibilities to participate in the evaluation is necessary.

➢ The initial budget should include a sufficient amount of resources to conduct the evaluation for an appropriate length of time.

➢ The data collecting, entering, and analysis functions should be identified and be adequate to ensure accurate data reporting.

➢ A plan for how, when, and to whom data will be reported is also critical.

EXAMPLES OF SMALL-SCALE AND LARGE-SCALE PROGRAM EVALUATIONS:

Rhode Island
Rhode Island conducted an evaluation of their pilot project to implement early learning guidelines. The pilot conducted in 2002 was completed by an outside evaluator to assess both the effectiveness of the training and to validate the relevance and use of the early learning guidelines document. As a result of the evaluation, revisions were made to the early learning standards and to the training content and delivery. For more information, contact Barbara Burgess at bburgess@ride.ri.net or Reeva Sullivan Murphy at rmurphy@dhs.ri.gov.

North Carolina
A study of 110 publicly funded preschool child care programs that participated in the Smart Start program from 1994 to 2001 measured the quality of classroom practice; participation levels in technical assistance activities; and a sample of children's abilities in language, literacy, numeracy, and social emotional skills. Results revealed that the quality of the centers that participated in the program steadily increased. Participation in technical assistance was significantly related to quality of the center, and children in higher-quality centers scored higher on readiness measures. Information is available on the Web at http://www.fpg.unc.edu/smartstart/.

The Build Initiative
The Build Initiative is a nine-State, multiyear initiative supported by a number of the foundations that participate in the Early Childhood Funders' Collaborative. The evaluation of the Build Initiative began with four States—Illinois, Minnesota, New Jersey, and Ohio—in May 2002. In 2003, Build added a fifth State, Pennsylvania, as part of the national evaluation; and four learning partner States—Hawaii, Michigan, Oklahoma, and Washington. A Build Initiative, national evaluation consultant helped each State identify a State evaluation partner. The State evaluation partner observes and participates in Build planning activities at the State level and is responsible for producing a site case study, which is later incorporated into a larger national report. Information is available on the Web at http://buildinitiative.org/.

RESOURCES:

For a free subscription and information on current issues facing program evaluators, including innovative methods, emerging trends, and approaches to evaluation, see the following:
- *The Evaluation Exchange*, published by the Harvard Family Research Project. This resource is available on the Web at http://www.gse.harvard.edu/hfrp/eval.html.

For further information on choosing assessment instruments for children to be used in program evaluation, see the following:
- "Early Childhood Assessment" (March 2003), an *Early Learning Policy Brief*, by Jessica McMaken, produced by Education Commission of the States. This resource is available on the Web at http://www.ecs.org/clearinghouse/43/19/4319.htm.

For a discussion of the lessons learned in Michigan's evaluation of the State preschool program, see the following:
- "Issues in Implementing a State Preschool Program Evaluation in Michigan" (2003), in *Assessing the State of State Assessments: Perspectives on Assessing Young Children,* by Lawrence Schweinhart, eds. Catherine Scott-Little, Sharon Lynn Kagan, and Richard M. Clifford, produced by the Southeastern Regional Vision for Education (SERVE). This resource is available on the Web at http://www.serve.org/_downloads/REL/ELO/ASSA.pdf.

For information on evaluations of school readiness programs, including a discussion of the methodologies used, effects on child development; and specific program features linked to effectiveness, see the following:
- *Evaluations of School Readiness Initiatives: What Are We Learning?* (March 2003), by Elizabeth Brown and Catherine Scott-Little, produced by the Southeastern Regional Vision for Education (SERVE). This resource is available on the Web http://www.serve.org/_downloads/REL/ELO/SchoolReadiness.pdf.

HOW ARE STATES EVALUATING QUALITY?

Many States have been determining if their use of the quality set-asides from the Child Care and Development Fund has been effective. States have also used a number of other strategies to improve the quality of child care, such as tiered reimbursement or rated licenses, and are increasingly engaged in evaluating the effects of these strategies. A growing body of research is supporting States' efforts to define and measure quality, although challenges remain due to the diversity of the field and the nature of assessing young children in dynamic environments.

KEY POINTS:

➢ Researchers define quality in two ways: structural and process. *Structural* quality refers primarily to aspects of care that can be regulated, such as licensing, adult-child rations, credentials, and education requirements. *Process* quality refers to environmental influences on quality, such as adult-child interaction, learning activities, and other characteristics related to caregivers' styles of relating to children. The process aspects of quality are much more difficult to measure.

➢ Currently, the most common tool used to measure process quality is an environmental or classroom rating scale. The most commonly used measure is the *Early Childhood Environmental Rating Scale (ECERS)*, including versions for infant-toddler, the Infant Toddler Environmental Rating Scale *(ITERS)* and for school age care, the *School Age Classroom Environmental Rating Scale (SACERS).*

➢ The *ECERS* requires classroom observations by trained observers who have achieved a specific level of reliability; although it is also used in some programs as a self-assessment tool.

➢ New tools are emerging, such as the *Early Language and Literacy Classroom Observation Tool (ELLCO)* that measures the degree to which a classroom and the teachers in the classroom support young children's literacy development.

There is an urgent need for more appropriate and valid tools for measuring quality in a variety of care settings. A new tool about to be released by Bank Street College is designed to assess the quality of relative care (see the reference under resources on page 7).

MEASURES COMMONLY USED IN EVALUATING QUALITY:

There are four Environment Rating Scales (ERS), each designed for a different segment of the early childhood field. Each one of the scales has items to evaluate: Physical Environment, Basic Care, Curriculum, Interaction, Schedule and Program Structure, and Parent and Staff Education.

The Early Childhood Environment Rating Scale, Revised Edition (ECERS-R) (1998), by Thelma Harms, Richard M. Clifford, and Debby Cryer, is designed to assess group programs for children of preschool through kindergarten age, 2½ through 5.

Escala de Calificación del Ambiente de la Infancia Temprana, Edición Revisada (2002), by Thelma Harms, Richard M. Clifford, and Debby Cryer, translated by Corina Dueñas, is the Spanish version of the *ECERS-R*.

The Infant/Toddler Environment Rating Scale, Revised Edition (ITERS-R) (2003), by Thelma Harms, Debby Cryer, and Richard M. Clifford, is designed to assess group programs for children from birth to 2½ years of age.

The Family Day Care Rating Scale (FDCRS) (1989), by Thelma Harms and Richard M. Clifford, is designed to assess family child care programs conducted in a provider's home.

The School-Age Care Environment Rating Scale (SACERS) (1995), by Thelma Harms, Ellen Vineberg Jacobs, and Donna Romano, is designed to assess group-care programs for school-age children 5 to 12 years of age.

Another commonly used tool is the *The Caregiver Interaction Scale* (1989), developed by Jeffrey Arnett, designed to assess the quality of caregivers' interaction with children.

RESOURCES:

For information on how to use results-based decision-making in designing evaluations for early care and education programs, particularly those involving multiple partners, see the following:
- *Using Results to Improve the Lives of Children and Families: A Guide for Public-Private Child Care Partnerships* (n.d.), by the Child Care Partnership Project. This resource is available on the Web at http://nccic.org/ccpartnerships/results.pdf

For a brief guide to evaluation and a set of 14 user-friendly sample instruments that can be used for both formative and summative evaluations of child care quality improvement initiatives, see the following:
- *A Toolkit for Evaluating Initiatives to Improve Child Care Quality* (2003), by Toni Porter, et al., produced by Bank Street College of Education, Institute for a Child Care Continuum. This resource is available on the Web at http://www.bankstreet.edu/iccc/toolkit.html

A related reference to the above Toolkit, which includes information on a variety of quality initiatives and their implementation and evaluation strategies, see the following:
- *Assessing Child Care Development Fund (CCDF) Investments In Child Care Quality: A Study of Selected State Initiatives* (2002), available on the Web at http://www.bankstreet.edu/iccc/publications.html.

Also check the Web site for the upcoming *Child Care Assessment Tool for Relatives*, which will be available on the Web at http://www.bankstreet.edu/iccc/.

For information on the research on quality and the policy implications of public funding of efforts to improve quality, see the following:
- *Child Care Quality: Does It Matter and Does It Need to be Improved?* (May 2000), by Deborah Vandell and Barbara Wolfe, produced by the Office of the Assistant Secretary for

Planning and Evaluation, U.S. Department of Health and Human Services. This resource is available on the Web at http://aspe.hhs.gov/hsp/ccquality00/index.htm.

For information on specific indicators of quality, measures used, and results found in recent research and an easy to read matrix of the recent research on the overall effects of child care quality, see the following:

- *Keeping Current in Child Care Research Annotated Bibliography: An Update (Spring 2002*) by Deborah Ceglowski & Chiara Bacigalupa, Early Childhood Research and Practice. This resource is available on the Web at http://aspe.hhs.gov/hsp/ccquality00/index.htm.

For information on specific measures used by States and researchers, see the following two Web sites:

- **Child Care and Early Education Research Connections.** A 50-State Data Tool allows users to create their own tables of State data; and an Instruments and Measures Section allows users to review measures that have been used in previous research, as well as search for specific measures by key words or phrases. This Web site can be accessed at http://childcareresearch.org/discover/index.jsp.

- **National Institute for Early Education Research (NIEER).** The Assessment Database is a resource for finding information on a variety of assessments and for identifying assessments appropriate for specific uses. Users can define specific terms to build their own "custom" database or review "pre-built" databases developed by NIEER. A variety of related links are also identified. This Web site can be accessed at http://nieer.org/assessment/.

The National Child Care Information Center does not endorse any organization, publication, or resource.

Updated June 2005